A crystal base for the genetic code

L. Frappat, P. Sorba

Laboratoire de Physique Théorique ENSLAPP, URA 1436,
École Normale Supérieure de Lyon and Université de Savoie, France

A. Sciarrino

Dipartimento di Scienze Fisiche, Università di Napoli "Federico II"
and I.N.F.N., Sezione di Napoli, Italy

Abstract

The quantum enveloping algebra $\mathcal{U}_q(sl(2) \oplus sl(2))$ in the limit $q \to 0$ is proposed as a symmetry algebra for the genetic code. In this approach the triplets of nucleotids or codons in the DNA chain are classified in crystal bases, tensor product of $\mathcal{U}_{q\to 0}(sl(2) \oplus sl(2))$ representations. Such a construction might be compared to the baryon classification from quark building blocks in elementary particles physics, one of the main differences standing in the property of a crystal base to provide a natural order in the state constituents, this order being crucial in the codon. Then an operator ensuring the correspondence codon/amino-acid can be constructed out of the above algebra. It will be called the reading operator, and be such that two codons relative to the same (resp. different) amino-acid(s) acquire the same (resp. different) eigenvalue(s).

Résumé

L'algèbre enveloppante quantique $\mathcal{U}_q(sl(2) \oplus sl(2))$ dans la limite $q \to 0$ est proposée comme algèbre de symétrie du code génétique. Dans cette approche, les triplets de nucléotides ou codons dans la chaine d'ADN sont classifiés dans des bases cristallines, produit tensoriel de représentations de $\mathcal{U}_{q\to 0}(sl(2) \oplus sl(2))$. Une telle construction peut être comparée à la classification des baryons à partir des quarks en physique des particules élémentaires, une des différences essentielles résidant dans la propriété d'une base cristalline de fournir un ordre naturel des constituents, cet ordre étant crucial dans le codon. Nous construisons un opérateur assurant la correspondance codon/acide aminé, appelé opérateur de lecture. Cet opérateur est tel que deux codons relatifs au même (resp. à des différents) acide(s) aminé(s) ont des valeurs propres identiques (resp. différentes).

ENSLAPP-AL-671/97
DSF-97/37
physics/9801027
December 1997

The mystery of the perfect correspondence between triplets of nucleotides or codons in the desoxyribonucleic acid (DNA) sequence and the amino-acids is called the genetic code [1]. Let us, in a few words, remind how the DNA conducts the synthesis of proteins, which constitute the most abundant organic substances in living matter systems. Indeed, the DNA macromolecule is made of two linear chains of nucleotides in the famous double helix structure. Each nucleotide is characterized by one of the four elementary bases: adenine (A) and guanine (G) deriving from purine, and cytosine (C) and thymine (T) coming from pyrimidine. The DNA is localized in the nucleus of the cell and the transmission of the genetic information in the cytoplasm is achieved by the messenger ribonucleic acid or mRNA. This operation is called the transcription, the A, G, C, T bases in the DNA being respectively associated to the U, C, G, A bases, U denoting the uracile base. Finally, a codon is an ordered sequence of three bases (e.g. AAG, ACG, etc.) and it is a simple exercise to numerate $4 \times 4 \times 4$ different codons. Except the three following triplets UAA, UAG and UGA, each of the 61 others is related through a ribosome to an amino-acid in the universal eukariotic code (see Table 2). Thus the chain of nucleotides in the mRNA – and also in the DNA – can also be viewed as a sequence of triplets, each corresponding to an amino-acid, except the three above mentioned ones. These last codons are called non-sense or stop-codons, and their role is to stop the biosynthesis.

One can distinguish 20 different amino-acids [1]. It follows that different codons can be associated to the same amino-acid, or in other words, that the genetic code is degenerated. Considering the standard eukariotic code (see Table 2), one remarks that the codons are organized in sextets, quadruplets, triplets, doublets and even singlets, each multiplet corresponding to a specific amino-acid. Such a picture naturally led Hornos and Hornos [2] to look for an underlying symmetry based on a continuous Lie group. More precisely, the authors tried to answer the following question: is it possible to determine a Lie group G carrying a 64-dimensional irreducible representation R and admitting a subgroup H such that the decomposition of R into irreducible representations under H gives exactly the different just above mentioned multiplets? They proposed as starting symmetry the symplectic group $Sp(6)$ with successive breakings up to its Cartan part $U(1) \times U(1) \times U(1)$.

Interpreting the double origin of the nucleotides, each arising either from purine or from pyrimidine, as a $\mathbb{Z}_2$-grading, a supersymmetric extension of the above model has been proposed [3] with the superalgebra $sl(6|1)$ as the classification (super)algebra before symmetry breaking. A systematic search for superalgebras the representation theory of which comes close to the multiplet structure of the genetic code has also been recently carried out in ref. [4].

It is a rather different point of view that we will adopt in this letter. Indeed we will consider the four nucleotides as basic states of the $(\frac{1}{2}, \frac{1}{2})$ representation of the $\mathcal{U}_q(sl(2) \oplus sl(2))$ quantum

[1] They are denoted by the letters Ala (Alanine), Arg (Arginine), Asn (Asparagine), Asp (Aspartic acid), Cys (Cysteine), Gln (Glutamine), Glu (Glutamic acid), Gly (Glycine), His (Histidine), Ile (Isoleucine), Leu (Leucine), Lys (Lysine), Met (Methionine), Phe (Phenylalanine), Pro (Proline), Ser (Serine), Thr (Threonine), Trp (Tryptophane), Tyr (Tyrosine), Val (Valine).

enveloping algebra in the limit $q \to 0$. Then a triplet of nucleotides will be obtained by constructing the tensor product of three such four dimensional representations. Actually, this approach mimicks the group theoretical classification of baryons made out from three quarks in elementary particles physics, the building blocks being here the A, C, G, T (U) nucleotides. The main and essential difference stands in the property of a codon to be an *ordered* set of three nucleotides, which is not the case for a baryon. Let us be more explicit on an example: there are three different codons made of the A, A, U nucleotides, namely AAU, AUA and UAA, while the proton appears as a weighted combination of the two u quarks and one d quark, that is $|p\rangle \sim |uud\rangle + |udu\rangle + |duu\rangle$, where the spin structure is implicit.

Constructing such pure states is made possible in the framework of any algebra $\mathcal{U}_{q\to 0}(\mathcal{G})$ with $\mathcal{G}$ being any (semi)-simple classical Lie algebra owing to the existence of a special basis, called crystal basis, in any (finite dimensional) representation of $\mathcal{G}$. The algebra $\mathcal{G} = su(2) \oplus su(2) \simeq so(4)$ appears the most natural for our purpose. First of all, it is "reasonable" to represent the four nucleotides in the fundamental representation of $\mathcal{G}$. Moreover, the complementary rule in the DNA–mRNA transcription may suggest to assign a *quantum number* with opposite values to the couples (A,T/U) and (C,G). The distinction between the purine bases (A,G) and the pyrimidine ones (C,T/U) can be algebraically represented in an analogous way. Thus considering the representation $(\frac{1}{2}, \frac{1}{2})$ of the group $SU(2) \times SU(2)$ and denoting $\pm$ the basis vector corresponding to the eigenvalues $\pm\frac{1}{2}$ of the J_3 generator in any of the two $su(2)$ corresponding algebras, we will assume the following "biological" spin structure:

$$\begin{array}{ccc}
 & su(2)_H & \\
C \equiv (+,+) & \longleftrightarrow & U \equiv (-,+) \\
 & & \\
su(2)_V \updownarrow & & \updownarrow su(2)_V \\
 & & \\
G \equiv (+,-) & \longleftrightarrow & A \equiv (-,-) \\
 & su(2)_H &
\end{array} \qquad (1)$$

the subscripts H (:= horizontal) and V (:= vertical) being just added to specify the group actions.

Now, let us turn our attention towards the representations of $\mathcal{U}_{q\to 0}(\mathcal{G})$ and more specifically to their crystal bases. In statistical mechanics, the $q \to 0$ limit of a deformed (quantum) algebra can be interpreted as the absolute zero temperature in a lattice model. Introducing in $\mathcal{U}_{q\to 0}(\mathcal{G})$ the operators $\tilde{e}_i$ and $\tilde{f}_i$ ($i = 1, \ldots, \mathrm{rank}\ \mathcal{G}$) after modification of the simple root vectors e_i and f_i of $\mathcal{U}_q(\mathcal{G})$, a particular kind of basis in a $\mathcal{U}_q(\mathcal{G})$-module can be defined. Such a basis is called a crystal basis and carries the property to undergo in a specially simple way the action of the $\tilde{e}_i$ and $\tilde{f}_i$ operators:

as an example, for any couple of vectors u, v in the crystal basis $\mathcal{B}$, one gets $u = \tilde{e}_i v$ if and only if $v = \tilde{f}_i u$. One must note that there is no objection to consider the four states C, U, G, A defined in (2) as constituting a crystal basis for the $(\frac{1}{2}, \frac{1}{2})$ module of $\mathcal{U}_{q\to 0}(sl(2) \oplus sl(2))$. More interesting for our purpose is the crystal basis in the tensorial product of two representations. Then the following theorem holds [5]:

Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be the crystal bases of the M_1 and M_2 $\mathcal{U}_{q\to 0}(\mathcal{G})$-modules respectively. Then for $u \in \mathcal{B}_1$ and $v \in \mathcal{B}_2$, we have:

$$\tilde{f}_i(u \otimes v) = \begin{cases} \tilde{f}_i u \otimes v & \exists\, n \geq 1 \text{ such that } \tilde{f}_i^n u \neq 0 \text{ and } \tilde{e}_i v = 0 \\ u \otimes \tilde{f}_i v & \text{otherwise} \end{cases} \tag{2}$$

$$\tilde{e}_i(u \otimes v) = \begin{cases} u \otimes \tilde{e}_i v & \exists\, n \geq 1 \text{ such that } \tilde{e}_i^n v \neq 0 \text{ and } \tilde{f}_i u = 0 \\ \tilde{e}_i u \otimes v & \text{otherwise} \end{cases} \tag{3}$$

To represent a codon, we will have to perform the tensor product of three $(\frac{1}{2}, \frac{1}{2})$ representations of $\mathcal{U}_{q\to 0}(sl(2) \oplus sl(2))$. However, it is well-known – and easy to check from Tables 1, 2 – that in a multiplet of codons relative to a specific amino-acid, the two first bases constituent of a codon are "relatively stable", the degeneracy being mainly generated by the third nucleotide. For that reason, we will prefer to examine a codon as a 2+1 state instead of a simple triplet. So, let us consider in detail the first tensor product:

$$(\tfrac{1}{2}, \tfrac{1}{2}) \otimes (\tfrac{1}{2}, \tfrac{1}{2}) = (1,1) \oplus (1,0) \oplus (0,1) \oplus (0,0) \tag{4}$$

where inside the parenthesis, $j = 0, \frac{1}{2}, 1$ is put in place of the $2j+1 = 1, 2, 3$ respectively dimensional $SU(2)$ representation. We get, using Theorem 1, the following tableau:

$$\begin{array}{llll} \rightarrow\ su(2)_H & (0,0) \quad (CA) & & (1,0) \quad (\ CG \quad UG \quad UA\ ) \\ \downarrow & & & \\ su(2)_V & (0,1) \begin{pmatrix} CU \\ GU \\ GA \end{pmatrix} & & (1,1) \begin{pmatrix} CC & UC & UU \\ GC & AC & AU \\ GG & AG & AA \end{pmatrix} \end{array}$$

From Tables 1 and 2, the dinucleotide states formed by the first two nucleotides in a codon can be put in correspondence with quadruplets, doublets or singlets of codons relative to an amino-acid. Note that the sextets (resp. triplets) are viewed as the sum of a quadruplet and a doublet (resp. a doublet and a singlet). The dinucleotide states associated to the quadruplets (as well as those included in the sextets) of codons satisfy:

$$J^d_{H,3} > 0 \quad \text{or} \quad J^d_{H,3} = 0\,, \quad J^d_{V,3} \geq 0\,, \quad J^d_V \neq 0\,. \tag{5}$$

where $J^d_{H,3}$ and $J^d_{V,3}$ are the third components of the spin generators of the dinucleotide states.
The dinucleotide states associated to the doublets (as well as those included in the triplets) and eventually to the singlets of codons are such that:

$$J^d_{H,3} < 0 \quad \text{or} \quad J^d_{H,3} = 0\,, \quad J^d_{V,3} < 0 \text{ or } J^d_V = 0\,. \tag{6}$$

On the other hand, if we consider the three-fold tensor product, the content into irreducible representations of $\mathcal{U}_{q\to 0}(sl(2)\oplus sl(2))$ is given by:

$$(\tfrac{1}{2},\tfrac{1}{2})\otimes(\tfrac{1}{2},\tfrac{1}{2})\otimes(\tfrac{1}{2},\tfrac{1}{2}) = (\tfrac{3}{2},\tfrac{3}{2})\oplus 2\,(\tfrac{3}{2},\tfrac{1}{2})\oplus 2\,(\tfrac{1}{2},\tfrac{3}{2})\oplus 4\,(\tfrac{1}{2},\tfrac{1}{2}) \tag{7}$$

The structure of the irreducible representations of the r.h.s. of Eq. (7) is:

$$(\tfrac{3}{2},\tfrac{3}{2}) \equiv \begin{pmatrix} CCC & UCC & UUC & UUU \\ GCC & ACC & AUC & AUU \\ GGC & AGC & AAC & AAU \\ GGG & AGG & AAG & AAA \end{pmatrix}$$

$$(\tfrac{3}{2},\tfrac{1}{2}) \equiv \begin{pmatrix} CCG & UCG & UUG & UUA \\ GCG & ACG & AUG & AUA \end{pmatrix}$$

$$(\tfrac{3}{2},\tfrac{1}{2})' \equiv \begin{pmatrix} CGC & UGC & UAC & UAU \\ CGG & UGG & UAG & UAA \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{3}{2}) \equiv \begin{pmatrix} CCU & UCU \\ GCU & ACU \\ GGU & AGU \\ GGA & AGA \end{pmatrix} \qquad (\tfrac{1}{2},\tfrac{3}{2})' \equiv \begin{pmatrix} CUC & CUU \\ GUC & GUU \\ GAC & GAU \\ GAG & GAA \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2}) \equiv \begin{pmatrix} CCA & UCA \\ GCA & ACA \end{pmatrix} \qquad (\tfrac{1}{2},\tfrac{1}{2})' \equiv \begin{pmatrix} CGU & UGU \\ CGA & UGA \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2})'' \equiv \begin{pmatrix} CUG & CUA \\ GUG & GUA \end{pmatrix} \qquad (\tfrac{1}{2},\tfrac{1}{2})''' \equiv \begin{pmatrix} CAC & CAU \\ CAG & CAA \end{pmatrix}$$

As expected from formulae (5) and (6), our model cannot gather codons associated to one particular amino-acid in the same irreducible multiplet. However, it is possible to construct an operator $\mathcal{R}$ out of the algebra $\mathcal{U}_{q\to 0}(sl(2)\oplus sl(2))$, acting on the codons, that will describe the genetic code in the following way:

Two codons have the same eigenvalue under $\mathcal{R}$ if and only if they are associated to the same amino-acid.

This operator will be called the *reading operator*. It has the following form:

$$\begin{aligned} \mathcal{R} \;=\;& \tfrac{4}{3}c_1\,C_H + \tfrac{4}{3}c_2\,C_V - 4c_1\,\mathcal{P}_1\,J_{H,3} - 4c_2\,\mathcal{P}_2\,J_{V,3} + (\mathcal{P}_3\,c_3 + \mathcal{P}_4\,c_4)\,J_{V,3} \\ & + \mathcal{P}_5\,c_5\,(\tfrac{1}{2} - J^{(3)}_{V,3}) + (\mathcal{P}_6\,q + \mathcal{P}'_6\,q')\,(\tfrac{1}{2} - J^{(3)}_{V,3})\,J^{(3)}_{H,3}\,. \end{aligned} \tag{8}$$

In Eq. (8), the operators $J_{H,3}$ and $J_{V,3}$ are the third components of the total spin generators of the algebra $\mathcal{U}_{q\to 0}(sl(2)\oplus sl(2))$, $J^{(3)}_{H,3}$, $J^{(3)}_{V,3}$ are the third components corresponding to the third nucleotide of a codon. Of course, these last two operators can be replaced by $J^{(3)}_{\alpha,3} = J_{\alpha,3} - J^{d}_{\alpha,3}$ ($\alpha = H, V$). The operator C_α ($\alpha = H, V$) is a "Casimir" operator of $\mathcal{U}_{q\to 0}(sl(2))$ in the crystal basis. It is

characterized by the property that it commutes with $J_{\pm,H}, J_{\pm,V}$ and $J_{H,3}, J_{V,3}$ (where $J_{\pm,H}, J_{\pm,V}$ are the generators with a well-defined behaviour for $q \to 0$) and its eigenvalues on any vector basis of an irreducible representation of highest weight J is $J(J+1)$, i.e. the same as the undeformed standard second degree Casimir operator of $sl(2)$. Its explicit expression is

$$C = (J_3)^2 + \frac{1}{2} \sum_{n \in \mathbb{Z}_+} \sum_{k=0}^{n} (J_-)^{n-k} (J_+)^n (J_-)^k \,. \tag{9}$$

Note that for $sl(2)_{q\to 0}$ the "Casimir" operator is an infinite series of powers of J_- and J_+. However in any finite irreducible representation only a finite number of terms gives a non-vanishing contribution. $\mathcal{P}_i$ $(i = 1, \ldots, 5)$ are projectors given by the following expressions:

$$\begin{aligned}
\mathcal{P}_1 &= J^d_{H+}\, J^d_{H-} \,, \\
\mathcal{P}_2 &= J^d_{V+}\, J^d_{V-} \,, \\
\mathcal{P}_3 &= J^d_{H-}\, J^d_{H+} (2 - J^d_{H+}\, J^d_{H-} - J^d_{V+}\, J^d_{V-}) + (1 - J^d_{H-}\, J^d_{H+})(1 - J^d_{H+}\, J^d_{H-})(1 - J^d_{V+}\, J^d_{V-}) \,, \\
\mathcal{P}_4 &= (J^d_{H-}\, J^d_{H+}) \left[(J^d_{H+}\, J^d_{H-})(1 - J^d_{V+}\, J^d_{V-}) + (J^d_{V+}\, J^d_{V-})(J^d_{V-}\, J^d_{V+})(1 - J^d_{H+}\, J^d_{H-}) \right] , \\
\mathcal{P}_5 &= (J^d_{H-}\, J^d_{H+})(J^d_{V-}\, J^d_{V+})(J^d_{H+}\, J^d_{H-})(1 - J^d_{V+}\, J^d_{V-}) \,.
\end{aligned} \tag{10}$$

The projectors $\mathcal{P}_6, \mathcal{P}'_6$ appear only for the eukariotic code. Their expressions are given by:

$$\begin{aligned}
\mathcal{P}_6 &= (J^d_{H-}\, J^d_{H+})(J^d_{V-}\, J^d_{V+})(1 - J^d_{H+}\, J^d_{H-})(J^d_{V+}\, J^d_{V-}) \,, \\
\mathcal{P}'_6 &= (J^d_{H-}\, J^d_{H+})(1 - J^d_{V-}\, J^d_{V+})(J^d_{H+}\, J^d_{H-})(1 - J^d_{V+}\, J^d_{V-}) \,.
\end{aligned} \tag{11}$$

The terms in c_1 and c_2 are responsible for the structure in quadruplets (given essentially by the dinucleotide content). The terms in c_3 give rise to the splitting of the quadruplets into doublets. The terms in c_4 and c_5 lead to the sextets. Finally, the terms in q and q', that appear only in the eukariotic code, are responsible for the singlet and triplet structure.

Now, using the values of the quantum numbers J_H, J_V, $J_{H,3}$, $J_{V,3}$, $J^d_{\alpha\pm}\, J^d_{\alpha\pm}$ $(\alpha = H, V)$ of the codons given in Tables 3 and 4, one can compute the action of the reading operator $\mathcal{R}$ on each of the 64 codons.

Although the eukariotic code (EC) seems to be a universal genetic code, it appears in some way as an advanced form of the vertebral mitochondrial code (VMC). Indeed there is very few difference between the two codes. The codons in the VMC are organized into 2 sextets, 6 quadruplets and 14 doublets. When evolving from the VMC to the EC, one doublet and one quartet merge together to form a sextet while two other doublets split into four singlets, two of them gluing with existing doublets to form two triplets. The final result for the EC is 3 sextets, 5 quadruplets, 10 doublets, 2 triplets and 2 singlets. Hence, it appears natural to start to calculate $\mathcal{R}$ for the vertebral mitochondrial code.

a) Vertebral Mitochondrial Code:

One finds the following eigenvalues of the reading operator $\mathcal{R}$ in the case of the vertebral mitochondrial code, identifying the amino-acids with its corresponding codons (Ser corresponds to the codons UCX (X=C,U,G,A) while Ser′ corresponds to the codons AGC/AGU; similarly Leu is related to the quartet CUX and Leu′ to the doublet UUG/UUA; finally, Arg is given by the quartet CGX and Ter′ to the doublet AGG/AGA):

$$
\begin{array}{ll}
\text{Pro} = -c_1 - c_2 & \text{Thr} = 3c_1 + 3c_2 \\
\text{Ala} = -c_1 + 3c_2 & \text{Ser} = 3c_1 - c_2 \\
\text{Asp} = c_1 + 5c_2 - \frac{1}{2}c_3 & \text{Glu} = c_1 + 5c_2 - \frac{3}{2}3c_3 \\
\text{Tyr} = 5c_1 + c_2 + \frac{1}{2}c_3 & \text{Ter} = 5c_1 + c_2 - \frac{1}{2}c_3 \\
\text{Asn} = 5c_1 + 5c_2 - \frac{1}{2}c_3 & \text{Lys} = 5c_1 + 5c_2 - \frac{3}{2}c_3 \\
\text{His} = c_1 + c_2 + \frac{1}{2}c_3 & \text{Gln} = c_1 + c_2 - \frac{1}{2}c_3 \\
\text{Arg} = -c_1 + c_2 & \text{Gly} = -c_1 + 5c_2 \\
\text{Cys} = 3c_1 + c_2 + \frac{1}{2}c_3 + \frac{1}{2}c_4 & \text{Trp} = 3c_1 + c_2 - \frac{1}{2}c_3 - \frac{1}{2}c_4 \\
\text{Ser}' = 3c_1 + 5c_2 - \frac{1}{2}c_3 - \frac{1}{2}c_4 & \text{Ter}' = 3c_1 + 5c_2 - \frac{3}{2}c_3 - \frac{3}{2}c_4 + c_5 \\
\text{Val} = c_1 + 3c_2 & \text{Leu} = c_1 - c_2 \\
\text{Phe} = 5c_1 - c_2 + \frac{3}{2}c_3 & \text{Leu}' = 5c_1 - c_2 + \frac{1}{2}c_3 \\
\text{Ile} = 5c_1 + 3c_2 + \frac{1}{2}c_3 + \frac{1}{2}c_4 & \text{Met} = 5c_1 + 3c_2 - \frac{1}{2}c_3 - \frac{1}{2}c_4
\end{array}
\tag{12}
$$

The parameters c_3, c_4 are fixed by the following requirements. The condition Leu = Leu′ leads to the expression of the coefficient c_3 in function of c_1 and c_2, and one obtains $c_3 = -8c_1$. At this point, one is led to add a correcting term in $\mathcal{R}$ since the symmetry of the genetic code implies Ile = Val and Cys = Arg as soon as Leu = Leu′ while Ser′ is not equal to Ser. Hence the projector $\mathcal{P}_4$ has a non-vanishing value on the AG, UG and AU dinucleotides. The condition Ser′ = Ser then implies $c_4 = 8c_1 + 12c_2$. At this point, Ile and Val on the one hand, and Cys and Arg on the other hand become different as required. Finally, the parameter c_5 is fixed for the VMC by requiring that Ter′ = Ter. One finds $c_5 = 6c_1 + 14c_2$. The demand to be satisfied by $\mathcal{R}$ in order to provide different eigenvalues to codons associated to different amino-acids implies the non-vanishing of c_1 and c_2. This leads, after a rescaling, to express the reading operator for the vertebral mitochondrial code as (where $c \equiv c_1/c_2$):

$$
\begin{aligned}
\mathcal{R}_{VMC}(c) \;=\; & \tfrac{4}{3}c\, C_H + \tfrac{4}{3}\, C_V - 4c\,\mathcal{P}_1\, J_{H,3} - 4\,\mathcal{P}_2\, J_{V,3} + (-8c\,\mathcal{P}_3 + (8c+12)\,\mathcal{P}_4)\, J_{V,3} \\
& + (6c+14)\,\mathcal{P}_5\,(\tfrac{1}{2} - J^{(3)}_{V,3})\,.
\end{aligned}
\tag{13}
$$

and therefore to the following values for the amino-acids:

a.a.	value of the codon	a.a.	value of the codon	a.a.	value of the codon
Ala	$-c+3$	Gly	$-c+5$	Pro	$-c-1$
Arg	$-c+1$	His	$-3c+1$	Ser	$3c-1$
Asn	$9c+5$	Ile	$5c+9$	Thr	$3c+3$
Asp	$5c+5$	Leu	$c-1$	Trp	$3c-5$
Cys	$3c+7$	Lys	$17c+5$	Tyr	$c+1$
Gln	$5c+1$	Met	$5c-3$	Val	$c+3$
Glu	$13c+5$	Phe	$-7c-1$	Ter	$9c+1$

(14)

The vertebral mitochondrial code

We remark that the reading operator $\mathcal{R}_{VMC}(c)$ can be used for any real value of c, except those confering the same eigenvalue to codons relative to two different amino-acids. These forbidden values are the following: -7, -5, -4, -3, $-\frac{5}{2}$, $-\frac{7}{3}$, -2, $-\frac{5}{3}$, $-\frac{3}{2}$, $-\frac{4}{3}$, -1, $-\frac{5}{6}$, $-\frac{4}{5}$, $-\frac{3}{4}$, $-\frac{5}{7}$, $-\frac{2}{3}$, $-\frac{3}{5}$, $-\frac{1}{2}$, $-\frac{3}{7}$, $-\frac{2}{5}$, $-\frac{3}{8}$, $-\frac{1}{3}$, $-\frac{3}{10}$, $-\frac{2}{7}$, $-\frac{1}{4}$, $-\frac{2}{9}$, $-\frac{1}{5}$, $-\frac{1}{6}$, $-\frac{1}{7}$, $-\frac{1}{8}$, $-\frac{1}{9}$, 0, $\frac{1}{7}$, $\frac{1}{6}$, $\frac{1}{5}$, $\frac{1}{4}$, $\frac{1}{3}$, $\frac{2}{5}$, $\frac{1}{2}$, $\frac{2}{3}$, 1, $\frac{4}{3}$, $\frac{3}{2}$, 2, $\frac{5}{2}$, 3, 4, 5.

b) The Eukariotic Code:

In the case of the eukariotic code, most of the eigenvalues of the reading operator are the same. The difference between VMC and EC comes i) from the doublets Met and Trp that split into singlets Met (AUG) + Ile$''$ (AUA) and Trp (UGG) + Ter$''$ (UGA), and ii) from the doublet Ter$'$ that merge with the quartet Arg to form a sextet. The eigenvalues for the new structures are the following:

$$\begin{aligned}
\mathrm{Ter}'' &= 3c_1 + c_2 - \tfrac{1}{2}c_3 - \tfrac{1}{2}c_4 - q' \\
\mathrm{Trp} &= 3c_1 + c_2 - \tfrac{1}{2}c_3 - \tfrac{1}{2}c_4 + q' \\
\mathrm{Ile}'' &= 5c_1 + 3c_2 - \tfrac{1}{2}c_3 - \tfrac{1}{2}c_4 - q \\
\mathrm{Met} &= 5c_1 + 3c_2 - \tfrac{1}{2}c_3 - \tfrac{1}{2}c_4 + q
\end{aligned} \tag{15}$$

The parameters c_3, c_4 are given as in the VMC. The parameter c_5 is now fixed by the condition Ter$'$ = Arg. One obtains $c_5 = -4c_1 + 14c_2$. The parameters q and q' describe the splitting of the doublets Met and Trp into the singlets: they are determined by requiring Ile$''$ = Ile and Ter$''$ = Ter. Il follows that $q = -12c_2$ and $q' = -6c_1 - 6c_2$. Hence the reading operator for the eukariotic code reads as:

$$\begin{aligned}
\mathcal{R}_{EC}(c) \;=\;& \tfrac{4}{3}c\,C_H + \tfrac{4}{3}\,C_V - 4c\,\mathcal{P}_1\,J_{H,3} - 4\,\mathcal{P}_2\,J_{V,3} + (-8c\,\mathcal{P}_3 + (8c+12)\,\mathcal{P}_4)\,J_{V,3} \\
&+(-4c+14)\,\mathcal{P}_5\,(\tfrac{1}{2} - J_{V,3}^{(3)}) - 6\,(2\,\mathcal{P}_6 + (c+1)\,\mathcal{P}'_6)\,(\tfrac{1}{2} - J_{V,3}^{(3)})\,J_{H,3}^{(3)}\,.
\end{aligned} \tag{16}$$

where as in case (a) we have achieved a rescaling and $c \equiv c_1/c_2$. This leads to the following values

for the amino-acids:

a.a.	value of the codon	a.a.	value of the codon	a.a.	value of the codon
Ala	$-c+3$	Gly	$-c+5$	Pro	$-c-1$
Arg	$-c+1$	His	$-3c+1$	Ser	$3c-1$
Asn	$9c+5$	Ile	$5c+9$	Thr	$3c+3$
Asp	$5c+5$	Leu	$c-1$	Trp	$-3c-11$
Cys	$3c+7$	Lys	$17c+5$	Tyr	$c+1$
Gln	$5c+1$	Met	$5c-15$	Val	$c+3$
Glu	$13c+5$	Phe	$-7c-1$	Ter	$9c+1$

(17)

The eukariotic code

As in the case (a), we have to rule out the values of the parameter c such that different amino-acids get the same eigenvalue under $\mathcal{R}_{EC}(c)$. The forbidden values now are the following: -8, -7, -6, -5, -4, $-\frac{7}{2}$, -3, $-\frac{5}{2}$, $-\frac{7}{3}$, -2, $-\frac{5}{3}$, $-\frac{3}{2}$, $-\frac{4}{3}$, -1, $-\frac{5}{6}$, $-\frac{4}{5}$, $-\frac{3}{4}$, $-\frac{2}{3}$, $-\frac{3}{5}$, $-\frac{1}{2}$, $-\frac{3}{7}$, $-\frac{2}{5}$, $-\frac{3}{8}$, $-\frac{1}{3}$, $-\frac{3}{10}$, $-\frac{2}{7}$, $-\frac{1}{4}$, $-\frac{2}{9}$, $-\frac{1}{5}$, $-\frac{1}{6}$, $-\frac{1}{7}$, $-\frac{1}{8}$, $-\frac{1}{9}$, 0, $\frac{1}{7}$, $\frac{1}{5}$, $\frac{1}{4}$, $\frac{1}{3}$, $\frac{2}{5}$, $\frac{1}{2}$, $\frac{2}{3}$, 1, $\frac{7}{6}$, $\frac{3}{2}$, 2, $\frac{7}{3}$, $\frac{5}{2}$, $\frac{8}{3}$, 3, $\frac{10}{3}$, $\frac{7}{2}$, 4, $\frac{9}{2}$, 7, 9, 11.

The simple model that we propose needs obviously to be developed. First on the symmetry point of view, it would be nice to understand or at least to include naturally in our approach the existence of sextets. Concerning the group structure, we have chosen $\mathcal{U}_{q\to 0}(\mathcal{G})$ with a minimal group $G = SU(2) \times SU(2)$, keeping in mind a physical interpretation. Of course, a larger symmetry might be of some help. As a second step, it will be reasonable to consider a more realistic model including interactions among bases. We wish to be soon able to apply our approach on the one hand for mutations and on the other hand in the fundamental problem of genome sequence.

Let us end this note by the following general remark. There are intense efforts these days to develop an interface between physics and biology. Different approaches are considered, among them the study of the DNA as an ideal polymer in the framework of statistical physics. But no direct connection between biology and elementary particle physics already showed up, in our knowledge. We hope that our proposal will raise up the interest of elementary particle physics in biology.

Acknowledgements

One of us (P.S.) is indebted to G. Brun and J.L. Darlix, biologists in ENS Lyon, for providing with informations and for encouragements, to A. Figureau for useful discussions. The authors wish also to thank D. Arnaudon for mentioning to us ref. [5].

CCC	Pro	UCC	Ser	GCC	Ala	ACC	Thr
CCU	Pro	UCU	Ser	GCU	Ala	ACU	Thr
CCG	Pro	UCG	Ser	GCG	Ala	ACG	Thr
CCA	Pro	UCA	Ser	GCA	Ala	ACA	Thr
CUC	Leu	UUC	Phe	GUC	Val	AUC	Ile
CUU	Leu	UUU	Phe	GUU	Val	AUU	Ile
CUG	Leu	UUG	Leu	GUG	Val	AUG	Met
CUA	Leu	UUA	Leu	GUA	Val	AUA	Met
CGC	Arg	UGC	Cys	GGC	Gly	AGC	Ser
CGU	Arg	UGU	Cys	GGU	Gly	AGU	Ser
CGG	Arg	UGG	Trp	GGG	Gly	AGG	Ter
CGA	Arg	UGA	Trp	GGA	Gly	AGA	Ter
CAC	His	UAC	Tyr	GAC	Asp	AAC	Asn
CAU	His	UAU	Tyr	GAU	Asp	AAU	Asn
CAG	Gln	UAG	Ter	GAG	Glu	AAG	Lys
CAA	Gln	UAA	Ter	GAA	Glu	AAA	Lys

Table 1: The vertebral mitochondrial code.

CCC	Pro	UCC	Ser	GCC	Ala	ACC	Thr
CCU	Pro	UCU	Ser	GCU	Ala	ACU	Thr
CCG	Pro	UCG	Ser	GCG	Ala	ACG	Thr
CCA	Pro	UCA	Ser	GCA	Ala	ACA	Thr
CUC	Leu	UUC	Phe	GUC	Val	AUC	Ile
CUU	Leu	UUU	Phe	GUU	Val	AUU	Ile
CUG	Leu	UUG	Leu	GUG	Val	AUG	Met
CUA	Leu	UUA	Leu	GUA	Val	AUA	Ile
CGC	Arg	UGC	Cys	GGC	Gly	AGC	Ser
CGU	Arg	UGU	Cys	GGU	Gly	AGU	Ser
CGG	Arg	UGG	Trp	GGG	Gly	AGG	Arg
CGA	Arg	UGA	Ter	GGA	Gly	AGA	Arg
CAC	His	UAC	Tyr	GAC	Asp	AAC	Asn
CAU	His	UAU	Tyr	GAU	Asp	AAU	Asn
CAG	Gln	UAG	Ter	GAG	Glu	AAG	Lys
CAA	Gln	UAA	Ter	GAA	Glu	AAA	Lys

Table 2: The eukariotic code.

codon	a.a.	J_H	J_V	$J_{H,3}$	$J_{V,3}$	codon	a.a.	J_H	J_V	$J_{H,3}$	$J_{V,3}$
CCC	Pro	3/2	3/2	3/2	3/2	GCC	Ala	3/2	3/2	3/2	1/2
CCU	Pro	1/2	3/2	1/2	3/2	GCU	Ala	1/2	3/2	1/2	1/2
CCG	Pro	3/2	1/2	3/2	1/2	GCG	Ala	3/2	1/2	3/2	-1/2
CCA	Pro	1/2	1/2	1/2	1/2	GCA	Ala	1/2	1/2	1/2	-1/2
CUC	Leu	1/2	3/2	1/2	3/2	GUC	Val	1/2	3/2	1/2	1/2
CUU	Leu	1/2	3/2	-1/2	3/2	GUU	Val	1/2	3/2	-1/2	1/2
CUG	Leu	1/2	1/2	1/2	1/2	GUG	Val	1/2	1/2	1/2	-1/2
CUA	Leu	1/2	1/2	-1/2	1/2	GUA	Val	1/2	1/2	-1/2	-1/2
CGC	Arg	3/2	1/2	3/2	1/2	GGC	Gly	3/2	3/2	3/2	-1/2
CGU	Arg	1/2	1/2	1/2	1/2	GGU	Gly	1/2	3/2	1/2	-1/2
CGG	Arg	3/2	1/2	3/2	-1/2	GGG	Gly	3/2	3/2	3/2	-3/2
CGA	Arg	1/2	1/2	1/2	-1/2	GGA	Gly	1/2	3/2	1/2	-3/2
CAC	His	1/2	1/2	1/2	1/2	GAC	Asp	1/2	3/2	1/2	-1/2
CAU	His	1/2	1/2	-1/2	1/2	GAU	Asp	1/2	3/2	-1/2	-1/2
CAG	Gln	1/2	1/2	1/2	-1/2	GAG	Glu	1/2	3/2	1/2	-3/2
CAA	Gln	1/2	1/2	-1/2	-1/2	GAA	Glu	1/2	3/2	-1/2	-3/2
UCC	Ser	3/2	3/2	1/2	3/2	ACC	Thr	3/2	3/2	1/2	1/2
UCU	Ser	1/2	3/2	-1/2	3/2	ACU	Thr	1/2	3/2	-1/2	1/2
UCG	Ser	3/2	1/2	1/2	1/2	ACG	Thr	3/2	1/2	1/2	-1/2
UCA	Ser	1/2	1/2	-1/2	1/2	ACA	Thr	1/2	1/2	-1/2	-1/2
UUC	Phe	3/2	3/2	-1/2	3/2	AUC	Ile	3/2	3/2	-1/2	1/2
UUU	Phe	3/2	3/2	-3/2	3/2	AUU	Ile	3/2	3/2	-3/2	1/2
UUG	Leu	3/2	1/2	-1/2	1/2	AUG	Met	3/2	1/2	-1/2	-1/2
UUA	Leu	3/2	1/2	-3/2	1/2	AUA	Met/Ile	3/2	1/2	-3/2	-1/2
UGC	Cys	3/2	1/2	1/2	1/2	AGC	Ser	3/2	3/2	1/2	-1/2
UGU	Cys	1/2	1/2	-1/2	1/2	AGU	Ser	1/2	3/2	-1/2	-1/2
UGG	Ter/Trp	3/2	1/2	1/2	-1/2	AGG	Ter/Arg	3/2	3/2	1/2	-3/2
UGA	Ter	1/2	1/2	-1/2	-1/2	AGA	Ter/Arg	1/2	3/2	-1/2	-3/2
UAC	Tyr	3/2	1/2	-1/2	1/2	AAC	Asn	3/2	3/2	-1/2	-1/2
UAU	Tyr	3/2	1/2	-3/2	1/2	AAU	Asn	3/2	3/2	-3/2	-1/2
UAG	Ter	3/2	1/2	-1/2	-1/2	AAG	Lys	3/2	3/2	-1/2	-3/2
UAA	Ter	3/2	1/2	-3/2	-1/2	AAA	Lys	3/2	3/2	-3/2	-3/2

Table 3: J_α, $(J_{\alpha,3})$: values of total/third component spin of $su(2)_\alpha$, $(\alpha = H, V)$.

(in the columns "amino-acids" (a.a.), left is for VMC and right for EC)

dinucl.	$J^d_{H+}\,J^d_{H-}$	$J^d_{H-}\,J^d_{H+}$	$J^d_{V+}\,J^d_{V-}$	$J^d_{V-}\,J^d_{V+}$
CC	1	0	1	0
CU	0	0	1	0
CG	1	0	0	0
CA	0	0	0	0
UC	1	1	1	0
UU	0	1	1	0
UG	1	1	0	0
UA	0	1	0	0
GC	1	0	1	1
GU	0	0	1	1
GG	1	0	0	1
GA	0	0	0	1
AC	1	1	1	1
AU	0	1	1	1
AG	1	1	0	1
AA	0	1	0	1

Table 4: Values of $J^d_{\alpha,\pm}\,J^d_{\alpha,\pm}$ $(\alpha = H, V)$ for the dinucleotides formed by the first two nucleotides.

References

[1] M. Singer, P. Berg, *Genes and Genomes,* Editions Vigot, Paris (1992).

[2] J.E. Hornos, Y. Hornos, *Algebraic model for the evolution of the genetic code,* Phys. Rev. Lett. **71** (1993) 4401.

[3] J.D. Bashford, I. Tsohantjis, P.D. Jarvis, *Supersymmetry and the evolution of the genetic code,* Preprint UTAS-PHYS-96-02 and *Codon and nucleotide assignments in a supersymmetric model of the genetic code,* Preprint UTAS-PHYS-96-45, to appear in Phys. Lett. A.

[4] S. Sachse, M. Forger, *Lie superalgebras and the multiplet structure of the genetic code,* Contribution to the 5th Int. Wigner Symposium, 1997, Vienne (Austria).

[5] M. Kashiwara, *Crystalizing the q-analogue of the universal enveloping algebra,* Commun. Math. Phys. **133** (1990) 249.